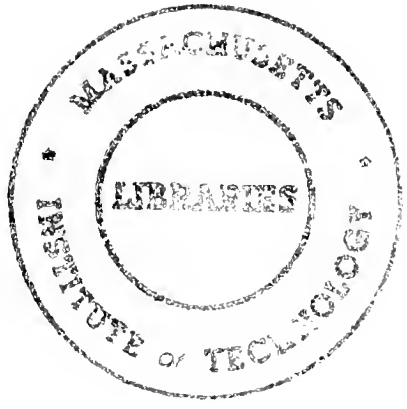

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
LIBRARIES

HD28
.M414
no.
3398-
92

TECHNICAL REPORT

Skip and Scan: Cleaning Up Telephone Interfaces

Paul Resnick
Robert A. Virsi

CCS TR# 129
SS WP# 3398-92

CENTER FOR
COORDINATION
SCIENCE

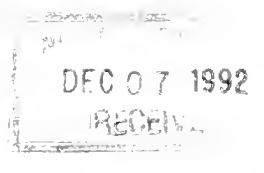
Massachusetts Institute of Technology
Sloan School of Management

Skip and Scan: Cleaning Up Telephone Interfaces

Paul Resnick
Robert A. Virsi

CCS TR# 129
SS WP# 3398-92

February, 1992


DEC 0 7 1992

Skip and Scan: Cleaning Up Telephone Interfaces

Paul Resnick
MIT Center for Coordination Science
Cambridge, MA 02139
presnick@eagle.mit.edu

Robert A.Virzi
GTE Laboratories Incorporated
Waltham, MA 02251
rv01@gte.com

ABSTRACT

The current generation of telephone interfaces is frustrating to use, in part because callers have to wait through the recitation of long prompts in order to find the options that interest them. In a visual medium, users would shift their gaze in order to skip uninteresting prompts and scan through large pieces of text. We present *skip and scan*, a new telephone interface style in which callers issue explicit commands to accomplish these same skipping and scanning activities. In a laboratory experiment, subjects made selections using skip and scan menus more quickly than using traditional, numbered menus, and preferred the skip and scan menus in subjective ratings. In a field test of a skip and scan interface, the general public successfully added and retrieved information without using any written instructions.

KEYWORDS: phone-based interface, semi-structure, audiotex, telephone form, menu, interactive voice response.

INTRODUCTION

Most people in the United States have used telephone information systems of some sort, many will admit that such systems are useful, but few people like to use them. With the current generation of telephone interfaces, callers are forced to wait through the recitation of long prompts and information when only selected pieces are of interest. We describe a new interface style, which we call *skip and scan*, that gives users more control over the process of listening and recording. In this style, the implicit structure of recorded prompts and information is made explicit and available to users for navigation purposes. Initial evidence indicates that the new style is preferred by users and lets them access information significantly faster. This may enable the creation of more complicated telephone-based information services, including groupware systems in which callers add information as well as retrieve it.

Reprinted from CHI '92: Human Factors in Computing Systems, Monterey, CA, May, 1992. Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

Hypermedia graphs are a convenient notation in which to describe telephone interfaces [2, 11, 12]. A graph determines what a caller will hear and what commands will be available during a telephone dialogue. A caller is always located at a particular node, and the sound associated with that node is played. Each node has links to other nodes, which are labelled by the commands a caller can use to traverse the links. The commands can be either touch-tone button presses, or verbal utterances entered via speech recognition. In addition, a default link may be traversed automatically after playing the sound for the current node, if no other command is entered. For example, the standard audiotex interface can be represented as a tree of nodes, with the sound for each node being prompts as to what buttons to press to follow links to other nodes (see Figure 1.) As we shall see later, the graph abstraction can also be used as an analytic tool, to relate user interface characteristics such as user control, consistency and simplicity to properties of graphs. We believe that an understanding of those graph properties will prove helpful in the design and evaluation of other new interaction techniques, both for telephones and for small-screen displays.

RELATED RESEARCH

Much work has gone into optimizing menus like those in Figure 1. There was some disagreement in the research community as to whether prompts should be presented in key-action order ("Press 1 for X") [4] or in action-key order ("For X, press 1") [3, 5], with the more recent research indicating that action-key is preferable. Most research indicates that three or four is the optimal number of options to be presented in a menu [3], though such advice is frequently not heeded since the categories that seem most natural often contain more than four items.

Two studies explored more unusual graph structures [13, 14]. Rosson and Mellen created a hierarchical graph in which each interior node contained a recording of a category name (e.g. entertainment, restaurants, or hotels.) Subjects were provided four buttons, two to move back and forth

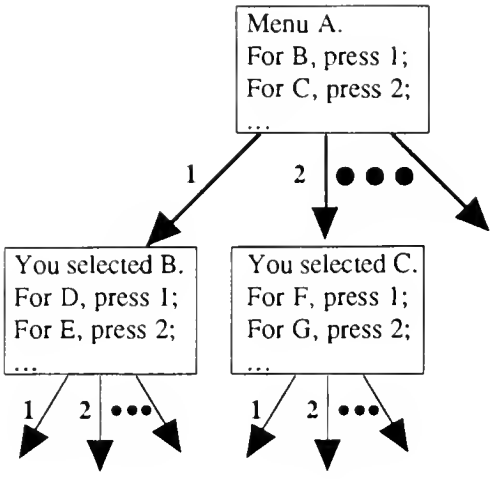


Figure 1: A generic audiotex system viewed as a hypermedia graph. Each node contains the prompts for one menu. Links are labeled by the touch-tone buttons that initiate link traversal.

between categories, one to select the current category, and one to move back up the hierarchy. Unfortunately, the mapping of information to the graph structure was not considered a variable in the study. Its novel features were not discussed, nor was it compared to the more conventional style of providing prompts for all of the categories in one node ("For entertainment, press 1; for restaurants, press 2; for hotels, press 3;...") Roberts and Engelbeck explored a spatial metaphor for navigation, in which information was laid out in a grid of nodes. The commands to navigate between nodes were spatially mapped to the telephone keypad (i.e., **2** up; **8** down; **4** left; **6** right.) They compared the spatial interface to a hierarchical menu interface, but found no significant differences in time required to perform tasks, or in subjective preferences. Our skip and scan information retrieval method builds on and generalizes the ideas in these two studies.

There is also much room for innovation in the way information is entered by telephone. Existing voice mail systems expect callers to leave an entire message as a single recording, thus leaving implicit any structure that the message might have. Most voice mail systems also begin recording when the system is ready rather than when callers are, which creates many awkward beginnings of messages. Recently, some voice mail systems have begun to give callers the option of reviewing and re-recording their messages.

The PhoneSlave conducted conversations with callers to elicit the several pieces of information it considered essential to good phone messages [15]. The system asked each caller a series of questions ("Who's calling please", "What is this in reference to?", "At what number can he reach you?", etc.) After playing a question, it recorded whatever the caller said, until a long pause was detected, then went on to the next question. While callers fill in the contents of a predefined structure with the PhoneSlave, Hindus [6] is exploring ways for participants in a phone conversation to add structure at their own discretion.

A NEW INTERFACE STYLE

We have developed a new interaction style for both information retrieval and entry. Information retrieval is still based on traversal of a menu hierarchy but the implicit structure of menus is made explicit. That enables users to skip and scan through the prompts within a menu. For information entry, users skip and scan through a series of separate but related entry blanks in a telephone form. Taken together, the retrieval and entry techniques are the basis of a coherent interface style that we call skip and scan. In this section, we first describe the retrieval technique and then the entry technique.

Retrieval: Skip and Scan Menus

The skip and scan style of selecting an option from a menu gives users more control over what prompts they hear. Figure 2 shows the skip and scan version of the top menu node from Figure 1. Each option described in the text of the original node becomes its own node in the new graph. Callers press **9** and **7** to skip forward and back between options and can always select the current option by pressing **1**. While the new menu style may look more cumbersome, it actually allows callers to scan through the options much more quickly, because they can skip ahead without listening to complete prompts. In the next section we present the results of an experiment that confirms this claim.

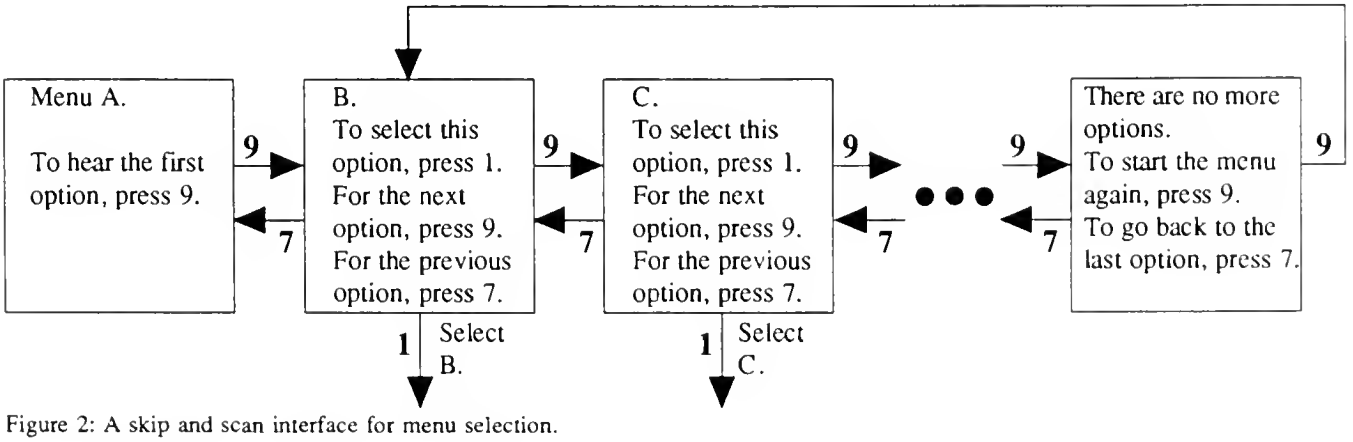


Figure 2: A skip and scan interface for menu selection.

Entry: Skip and Scan Forms

For information entry, we have developed the *telephone form* . We have generalized the PhoneSlave message taking dialogue to capture the structure of information objects other than personal phone messages. We have also turned from a conversational metaphor to a form-based metaphor. We believe that the form metaphor is more helpful, because it suggests an information entry process that is controlled by the user rather than suggesting an equal partnership between the user and a computer.

In a telephone form (Figure 3), there is one *entry blank* (node) for each separate recording that is expected. For example, if the object to be entered were an event announcement, there would be entry blanks for a headline, the date, the time, the location, etc. In addition to buttons **9** and **7** for navigating between entry blanks (note the consistent use of these buttons in Figures 2 and 3), the caller can use buttons **1** and **3** to record and/or erase the contents of the current entry blank. When the caller is satisfied with all of the recordings in the form, the caller presses # to save the entire object, or * to throw it away.

Semi-structured input has two advantages over making one long recording. First, the person recording is reminded of important information to include in the object, such as the admission price for an event. Second, splitting up an object into several separate recordings is a pre-requisite for allowing future callers to skip and scan through the logical segments of the object, a technique we will discuss in the future research section.

Our particular implementation of semi-structured input, the telephone form, provides users with a great deal of control over the entry process. Callers can quickly scan through the entry blanks to find out what information will be requested, so that they can better judge what information to record in each entry blank. They can gather their thoughts before starting each recording. In addition, users can recover from mistakes by re-recording single entry blanks rather than entire objects.

The telephone form concept can also be generalized to allow entry blanks that contain non-voice data. For example, an event announcement could contain a date entered using touch-tones. We have also implemented forms with entry blanks that contain links to other objects and lists of objects, which opens up new horizons for applications. Experience with visual interfaces has shown that lists of semi-structured objects, where the objects can contain links to other objects or lists, form the core of many group communications applications [8, 10].

EXPERIMENT: SELECTING A NAME FROM A LIST

We conducted an experiment that compared skip and scan menus with the more conventional menu style. Users were asked to find a target name from a list of between 3 and 12 names. Two methods of selection were tested in a within subjects design. In the standard method of selection, each name was announced followed by a selection number (Bob Smith, press 1; Paul Jones, press 2; etc.). This was compared to the skip and scan method as outlined in Figure 2, with one node for each name. Although we anticipated that users' overall performance would be faster in the skip and scan method, we expected to find evidence of a learning effect due to the novelty of the new method. We were interested in determining how many trials would be required before users performed as well with the skip and scan method as with the standard interface. We also asked users which style they preferred.

Methods

Subjects Two groups of subjects were run in this experiment. The first group was composed of 12 subjects recruited from a local university (mean age 23). We expected this group to perform well on the tasks and to exhibit relatively fast learning. To test the limits of applicability of this new technique, a second group of 6 subjects was drawn from an older population (mean age 62). We chose this older population because past experience has indicated that older users tend to be resistant to new technology and to have greater difficulty using telephone-based interfaces.

Stimuli A list of 100 names was randomly drawn from the telephone directory of a large corporation. Each name was

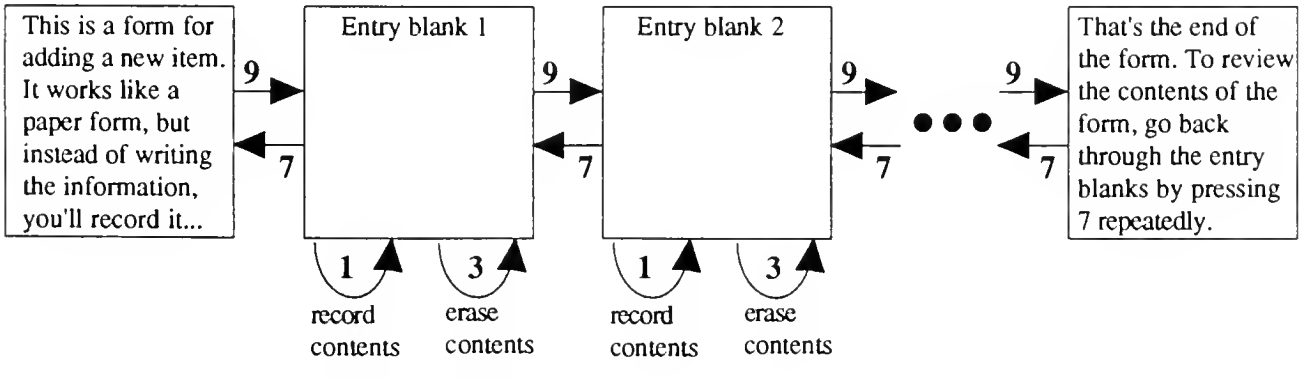


Figure 3: A telephone form containing several entry blanks.

presented as a first name followed by a last name exactly as it had appeared in the directory.

A total of 72 trials were prepared. Each trial consisted of a target name drawn randomly from the list of names and from 2 to 11 distractor names, leading to list lengths of 3 through 12 names. The target name appeared in each of the 12 serial positions 6 times. A random order was drawn for presenting the stimuli, and this same random order was used for both conditions and for all subjects.

A telephone interface was constructed that implemented each of the selection techniques. One female voice was used for all system prompts and a second female voice was used to present each of the names composing the lists. Users interacted with the systems from a telephone by pressing tone generating keys.

Procedures: Subjects were escorted into a testing room and seated before a standard desk set telephone. The general experimental procedures were explained but they were given absolutely no instruction on how they were to interact with the system. Instead they were told that they were to imagine that they had called a company with an automated directory service. They were told to follow the directions given by the system and to select the target name. Half the subjects in each group were presented the standard method first, the other half of the subjects interacted with the skip and scan method first. Between conditions they were warned that the method of selection had changed, and that they should attend to the instructions presented by the system.

Prior to the start of each trial, users listened to a name repeated over the telephone handset. This was the target name for the trial and it also appeared on a printed card next to the telephone as a memory aid. Users were told to press any key on the keypad when they were ready to begin the trial. Timing started when this key was pressed. In the skip and scan method, instructions, which the user had the option of skipping, were then played as part of a header node[1]. No instructions were required for the standard method. After each trial, users were told whether or not they had selected the correct name and then the next target name was announced.

After exposure to both systems, an overall preference question was asked, followed by an open-ended interview regarding the good and bad points of the two methods.

Results

Results are first presented for the group of 12 younger subjects, followed by the results for the 6 older subjects.

Younger Group: In Figure 4 the mean correct reaction times for the two conditions are shown as a function of target position. The best-fitting regression lines are superimposed. An Analysis of Variance (ANOVA) was calculated with the factors of Condition (standard menu method vs. skip and scan) and Target Position. The ANOVA confirms what the figure reveals. Overall, subjects were faster with the skip and scan method ($F(1,11) = 83.417$, $p<.001$) and were faster when the target name was earlier in the list ($F(11,121) = 140.572$, $p<.001$). Moreover, the interaction term was significant ($F(11,121) = 14.685$, $p<.001$) showing that the advantage for the skip and scan method is greater as the target name appears later in the list. This is not surprising as in the standard method subjects had to wait for the target item while in the skip and scan method users could jump forward in the list based on a match with the first name.

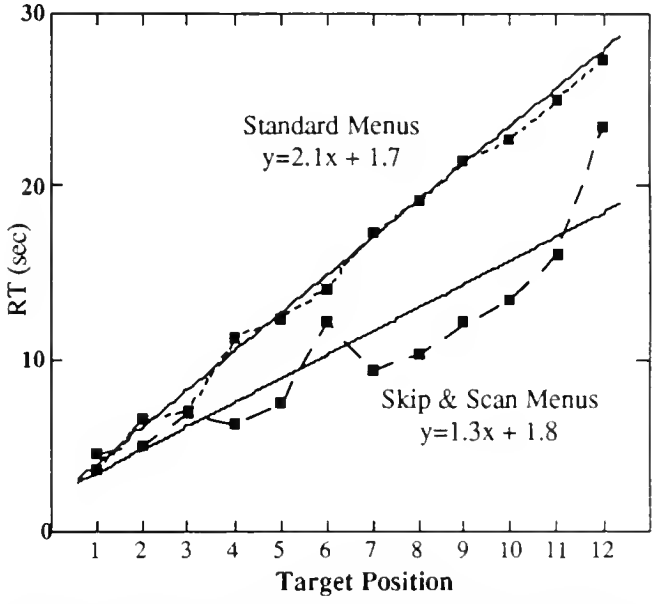


Figure 4. Mean correct reaction time is shown as a function of target position for the younger subject population. The regression equations appear next to each menu style.

We were interested in learning effects as well as overall performance. Because the trials were matched (i.e., the n^{th} trial in both conditions contained identical target names and lists of distractor names) we were able to calculate on a per subject basis two statistics that measure the learning effect. One statistic, the *crossover point*, was defined as the first trial on which the user was faster with the skip and scan interface. The other statistic, the *divergence point*, was defined as the beginning of the first run of five trials on which the user was faster on each trial with the skip and scan interface. The first statistic, the crossover point, had a mean value of 4.7 trials, a median of 3.0 trials, and a range of between 2 and 12 trials. The second statistic, the divergence point, had a mean value of 10.1 trials, a median of 6.0 trials, and a range of between 2 and 38 trials. Taken together, these results suggest that performance with the skip and scan menus surpassed performance with the skip and menus fairly rapidly.

[1] The instructions, if not interrupted, took 15 seconds to recite. The exact text was as follows: "<n> names are in the list. Scan through the names using 9 to skip ahead and 7 to skip backward. It's OK to interrupt the spoken voice at any time. Select a name by pressing 1. For the first name, press 9."

Error rates, although tracked, were too low to warrant analysis. In the skip and scan condition, errors were made on fewer than 1% of all trials. For the standard method, errors occurred on just over 2% of the trials. Most of these errors occurred on trials in which the user had to press two keys to make a selection (e.g., item number 10) when the second key was not pressed before the timeout so that the system interpreted the selection as item number 1.

When asked which system they preferred overall, all 12 subjects expressed a strong preference for the skip and scan method over the standard method ($p<.001$ by sign test). When probed as to why they preferred it, users stated that they thought it was faster, more efficient, and put them more in control.

Older Group: In Figure 5 the mean correct reaction times for the two conditions are shown as a function of target position with the regression lines superimposed. An ANOVA was performed with the factors of Condition (standard menu method vs. skip and scan) and Target Position. Unlike the younger subjects, the difference between the two methods was not reliable ($F(1,5) = 1.526$, $p>.10$). However, they were faster when the target name was earlier in the list ($F(11,55) = 59.492$, $p<.001$) and the interaction term was significant ($F(11,55) = 4.374$, $p<.001$). For this older population, the skip and scan method was slower when the target name was early in the list, but there was a small advantage when it later in the list.

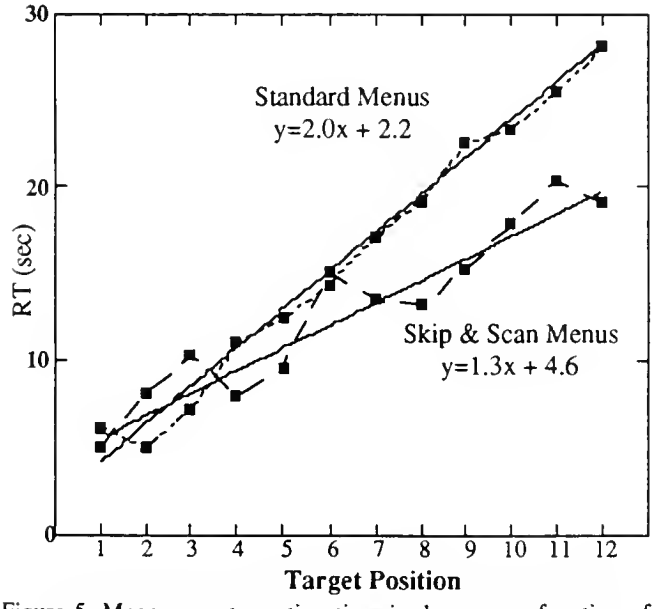


Figure 5. Mean correct reaction time is shown as a function of target position for the older subject population. The regression equations appear next to each menu style.

In general, the learning effect for this population was much more dramatic. As with the younger subjects, two statistics were calculated for each subject. The mean crossover point for the older subjects was 7.5 trials, the median was 12.5 trials, with a range of between 1 and 15 trials. The mean divergence point was at 21.0 trials, with a median of 13.5 trials, and a range of between 5 and 56 trials. When compared to the younger population, the older group clearly took longer to learn the new technique, primarily because of their resistance to interrupting the prompts.

Older subjects made considerably more errors than the younger subjects, with 9.6% and 10.4% errors for the skip and scan and standard methods, respectively. These high error rates are not surprising since the two interfaces were optimized for younger, faster subjects. Most of the incorrect trials on the skip and scan method occurred when subjects exceeded a threshold for time on the trial[2], often while asking questions of the experimenter. Most of the errors on the standard menus were of two kinds: subjects were too late in typing the second digit of two-digit selectors (e.g., the 2 of 12); or they associated the name with the number that preceded it rather than the one that followed it.

When asked which system they preferred, 5 of the 6 older subjects stated a preference for the skip and scan method ($p<.10$ by sign test). When asked for the reasons behind their preferences, all indicated that they preferred the one that seemed to be fastest.

Discussion

The results from the younger population strongly favor the skip and scan method. Not only did it lead to overall better performance, this benefit occurred within a few trials. We believe that the learning time might be reduced even further with a better wording of the introductory prompts that were in the header nodes of the skip and scan menus. The skip and scan method was also unanimously preferred to the standard method. Based on these results, it appears to us that the skip and scan menus have wide applicability in the development of voice response systems.

Because we wanted to test the limits of the technique, we ran an older group of subjects, drawn from a population known to be resistant to new technology. The clarity of the results with the younger population led us to believe that six subjects would be a sufficient sample from the older population.This population showed a marginally reliable preference for the skip and scan method, although their performance results were less clearly in favor of it. These subjects took longer to learn the new technique and their performance was better only on the longer menus. We would recommend caution in implementing the new technique to the extent that the user population was older, the system would have a significant number of first-time callers, or the menus tended to be naturally short.

COMPLETE APPLICATIONS

The experiment described above was motivated by two prototype applications, developed independently by the two

---

[2] Trials were counted as errors when the time to complete the trial was more than two standard deviations from that subject's mean, even if the subject selected the correct target.

authors of this paper. One was tested in the laboratory, the other in the field. Both were well received, although neither was compared against alternative telephone interfaces for the same application.

One application (developed by Virzi) was a community bulletin board containing newsclips about the activities of a town. The newsclips were grouped into categories. Category headings were presented as in Figure 2, and users traversed the list of categories using the **4** and **6** keys. Within each category, users traversed a list of up to 25 newsclips. Each newsclip consisted of a headline and contents. Users could traverse this list of headlines until they came upon an article of interest. Pressing the **2** key caused the contents of the article to be played.

Extensive usability testing (to be reported more completely elsewhere) indicated that there was some initial hesitancy and surprise that the system did not conform to users' expectations for a "standard" interface. However, performance was very good, with all users completing the tasks and navigating through the system within minutes.

The second application (developed by Resnick) was an events calendar used by Boston-area peace activists during and after the 1991 war with Iraq. After choosing a category, users could skip and scan through event announcements, sometimes as many as twenty in a category. Unlike the community bulletin board described above, the announcements were not split into separate headline and contents nodes. We hoped to achieve the same effect by having callers interrupt announcements after listening to the headline. Analysis of the keystroke logs for the 1973 calls handled between February 1 and April 5 indicates that they did: of the 1798 callers who pressed at least one touch-tone button, more than 90% interrupted at least one announcement.

The events hotline was very well-received by its users. The phone number was publicized through flyers at public rallies, and by word of mouth. The system sometimes handled more than 90 calls per day, which meant that the phone line was busy virtually all the time. Several individuals who claimed to be technophobes went out of their way to say that they liked it. Still, it is not clear how much of the positive response was due to the utility of the system (it had the most complete, up-to-date information about anti-war activities in Boston) rather than the usability of the interface. A month after the cease-fire, usage was down to 10-15 calls per day, where it has remained since then.

Perhaps most interestingly, the event announcements were added by the general public, using the telephone forms interface described in Figure 3. Each announcement consisted of six separate recordings for a headline, the date and time, the location, and so on. We estimate that at least 40 different people successfully filled out a form at one time or another. A few people had trouble with the concept of a form and added announcements that repeated the entire contents in several entry blanks.

The results from both of these applications are very encouraging. The skip and scan style allowed categories to contain 20 to 25 items. That, in turn, allowed users of the peace events hotline to add new items without necessitating frequent restructuring of the menus. Using conventional menus, with only a few items per menu, frequent restructuring would have been necessary.

The next version of the events hotline, about to be installed, separates the date and time into two entry blanks and prompts callers to type in a date rather than record it. That makes it possible to sort announcements by date and to throw out old announcements automatically. The new version also makes the structure of event announcements usable for information retrieval. Callers can skip back and forth between the segments of an announcement, making it possible, for example, to hear the dates and locations of all the event announcements in a category while skipping the other information.

GRAPH PROPERTIES

What makes it possible for users to skip and scan in a telephone interface? Above we used the hypermedia graph representation as a descriptive tool. Here we use it as an analytic tool: we operationalize the useful but vague principles of increased user control, consistency, and simplicity in terms of graph properties and the mappings of graphs to application concepts. We argue for graphs that have:

1) smaller nodes, each containing a headline;
2) a small number of consistently available links;
3) an easily recognizable mapping of graph nodes and links to application concepts.

Consider the differences between the graph styles of Figures 1 and 2. In Figure 1, the numbered-menu style, each node is larger, consisting of the prompts for all of the options that are available. Even if callers can reject options quickly, they have to wait through the remainder of the prompt for that option in order to hear the next prompt. In general, useful information should not follow useless information in the same node, but what is useful differs among users and contexts of usage. Smaller graph nodes, since they contain less information, are less likely to contain any out of order information.

Note that this argument in favor of smaller nodes does not apply to screen-based hypertext. In a screen-based environment, it is possible to present a large amount of text simultaneously and let users shift their eye gaze in order to skip and scan. The argument for smaller nodes would apply, however, to other "keyhole-sized" interfaces, such as very large print monitors and braille output devices that show only a few words at a time [7], and to the small LCD displays found on advanced-feature telephones, electronic address books, and the like.

Each node should contain a headline that describes the contents of the node. For example, in a telephone form,

each entry blank begins with a recitation of the name of the entry blank (e.g., "Date and time" or "Location"). The headline serves three functions. First, it provides an orientation cue as to where the user has arrived after following a link, which is important in visual hypermedia systems as well [9, 16]. Second, the headline summarizes the contents sufficiently to let callers decide whether it is safe to skip the node. Third, callers can get an overview of the contents of a group of nodes, by listening to just the headlines. Notice that the telephone form makes it possible to prompt callers to record a headline as a separate entry blank in new objects. In that way, the graph property of nodes having headlines can be maintained even as the general public adds nodes to a graph.

The ability of callers to control the dialogue depends not only on what actions they *can* take, but also on what actions they *know how* to take. Thus, callers may hear the headline for a node, know that there is something else they would rather be listening to, but not know how to skip to the other node. For example, Arons implemented a speech-only hypermedia graph with speech recognition used to specify which links to follow [2]. He found that his system spent more time prompting users about what links were available than playing the contents of the nodes (which also meant that callers had to listen through the nodes in order to hear the prompts.) He remedied this by making the links predictable, so that users would not always need to hear the prompts. For example, every node has a "more" link that goes to another node providing more details.

To make the links predictable, there should be both syntactic regularities in the graph and an easily understood mapping of the graph structure to the application structure. By syntactic regularities we mean that links with the same labels should be present at all or nearly every node. For example, in Figure 2, every node has a link labeled **9** emanating from it.

Even predicting that **9** is an available command will not be enough unless users can predict what it will do. Users can predict what **9** will do because there is a mapping between the nodes in the graph and the application concept of a set of options arranged in a list. Moreover, this mapping which is reinforced by the wording of the prompts. If callers understand the application concept and the mapping, they can predict that pressing **9** will move them to a node that contains a prompt for the next option in the list.

Contrast jumping between logical units of information, as just described, with jumping a fixed number of seconds, using fast-forward and rewind keys. With fast-forward and rewind, skipping is not coupled to the structure of the information, so users cannot be sure how much information they will be skipping, or what they will hear next. As an analogy, think about how hard it is to find a particular scene on a videotape using a VCR's fast-forward and rewind keys. Contrast this with a hypothetical system that pre-coded scene boundaries and allowed users to skip from scene to scene, always jumping to the beginning of a scene.

FUTURE RESEARCH

We are extending this research in several directions. First, the skip and scan style can be applied to additional tasks, particularly getting help and scanning the contents of a single object. Second, we plan additional experiments. Finally, we plan additional field tests, especially of groupware applications.

Even though our skip and scan menus contain just one "real" prompt per node, there are still "help" prompts for how to skip forward and back and how to select. If we apply the skip and scan style consistently, users should be able to skip through the help prompts to find the actions that interest them, without listening to the entire prompts for the other actions. We are trying different implementations of this idea, since the initial ones that we have tried make it harder for first-time callers to get started.

We are also exploring alternative ways to apply the skip and scan style to navigation through the entry blanks of objects that have been entered previously with telephone forms. One possibility is to use buttons **6** and **4** to skip between entry blanks of a single item, leaving **9** and **7** to skip between items. Another possibility is to play just a headline for each item. If a user selects an item, then **9** and **7** move between entry blanks until the user deselects the item. The latter implementation would maintain the simplicity of the interface but users would have to keep track of whether they were moving through a list of items or through the entry blanks of a single item.

We are designing additional laboratory experiments of different aspects of skip and scan interfaces. One experiment will explore the effects of numbering menu options as callers become more familiar with a graph and can remember the contents of some menus. In addition to the two item selection techniques described in this paper we will test a hybrid version that lets callers manually skip through the prompts in a menu, but still numbers the options, so that users can make a selection with a single keystroke if they do not need the prompts. We also plan an experiment to test our hypothesis that the ability to skip to a meaningful boundary in a long piece of speech is significantly more helpful than simple fast-forward and rewind keys.

The skip and scan interface style may broaden the scope of applications that can be successfully realized by telephone. For example, the telephone form metaphor may be a good vehicle for expressing queries by telephone. As another example, many group communication applications, such as meeting scheduling, organizational memory [1] and sales lead tracking, are just gaining acceptance in organizations in which everyone has access to networked computers. If the skip and scan paradigm makes it easier for callers to find information by telephone, such groupware applications might plausibly be implemented for telephones.

We have developed an application generator, called HyperVoice, to be reported elsewhere, that automatically

generates skip and scan interfaces, including the text of the prompts that need to be recorded. We are using HyperVoice to implement and, in some cases, test groupware applications. For example, we recently set up an organizational memory application for a group of teachers that lets them share questions, answers, and success stories related to a new curriculum project that they are participating in.

CONCLUSION

Skip and scan is a promising telephone interface style. Through explicit navigation commands, it gives users some of the control they get from shifting their gaze in visual interfaces. In a field trial, callers with no written instructions successfully used a skip and scan interface to both add and retrieve information. In an experiment, subjects preferred skip and scan menus to the more conventional, numbered menus. After an initial learning period, they also made selections more quickly with the skip and scan menus. The learning period was just a few trials for younger subjects and may be reduced with a more careful wording of prompts, or if skip and scan menus are used widely. Still, the need for a learning period of even one call may limit the utility of skip and scan menus in applications that are to be used predominantly by first-time callers.

ACKNOWLEDGEMENTS

We would like to thank Don Ottens for his able assistance in conducting the experiment reported in this paper. We would also like to thank Sandra Teare and Kate Dobroth for lending their voices, and Tom Lanning for helping to make this collaboration possible. Tom Malone contributed many important ideas throughout the course of this project, including the phrase "skip and scan." Mark Ackerman, Kum-Yew Lai, Jintae Lee, Brian Pentland, Mike Plusch, Troy Jordan, Debby Hindus and Chris Schmandt also contributed to the ideas and presentation of this paper.

REFERENCES

1. Ackerman, M. S. and Malone, T. W. Answer Garden: A Tool for Growing Organizational Memory. In *Conference on Office Information Systems*. (Cambridge, MA, 1990). ACM, pp. 31-39.

2. Arons, B. Hyperspeech: Navigating in Speech-Only Hypermedia. In Proceedings of *Hypertext*. (1991).

3. Engelbeck, G. and Roberts, T. The Effects of Several Voice-Menu Characteristics on Menu Selection Performance. Technical Report ST0401, US West Advanced Technologies, 1990.

4. Gould, J. D. and Boies, S. J. Human Factors Challenges in Creating a Principal Support Office System--The Speech Filing System Approach. *ACM Transactions on Office Information Systems*. 1, 4 (Oct. 1983), pp. 273-298.

5. Halstead-Nussloch, R. The Design of Phone-Based Interfaces for Consumers. In *CHI '89 Conference on Human Factors in Computing Systems*. (Austin, TX, 1989). ACM, pp. 347-352.

6. Hindus, D. Phone Calls Don't Have to Be Ephemeral: A Semi-Structured Approach to Stored Voice. Technical Report MIT Media Lab, 1991.

7. Ladner, R., Day, R., Gentry, D., Meyer, K. and Rose, S. A User Interface for Deaf-Blind People (Preliminary Report). In *CHI + GI '87*. (Toronto, 1987). ACM, pp. 75-80.

8. Lai, K.-Y., Malone, T. and Yu, K.-C. Object Lens: A "Spreadsheet" for Cooperative Work. *ACM Transactions on Office Information Systems*. 6, 4 (1988), pp. 332-353.

9. Landow, G. P. The Rhetoric of Hypermedia: Some Rules for Authors. *Journal of Computing in Higher Education*. 1, 1 (1989), pp. 39-64.

10. Malone, T. W., Grant, K. R., Lai, K.-Y., Rao, R. and Rosenblitt, D. Semi-structured Messages are Surprisingly Useful for Computer-Supported Coordination. In *Computer-Supported Cooperative Work: A Book of Readings*, Volume I. Greif, Ed. Morgan Kaufmann, San Mateo, CA, 1988, pp. 311-331.

11. Muller, M. J. and Daniel, J. E. Toward a Definition of Voice Documents. In *Proceedings of ACM Conference on Office Information Systems* (Boston, 1990) pp. 174-183.

12. Richards, J. T., Boies, S. J. and Gould, J. D. Rapid Prototyping and System Development: Examination of an Interface Toolkit for Voice and Telephony Applications. In *1986 Conference on Human Factors in Computing Systems*. (New York, 1986). ACM, pp. 216-220.

13. Roberts, T. L. and Engelbeck, G. The Effects of Device Technology on the Usability of Advanced Telephone Functions. In *CHI '89 Conference on Human Factors in Computing Systems*. (Austin, TX, 1989). ACM, pp. 331-337.

14. Rosson, M. B. and Mellen, N. M. Behavioral Issues in Speech-Based Remote Information Retrieval. In *AVIOS*. (San Francisco, CA, 1985).

15. Schmandt, C. and Arons, B. A Conversational Telephone Messaging System. *IEEE Transactions on Consumer Electronics*. CE-30, (1984).

16. Utting, K. and Yankelovich, N. Context and Orientation in Hypermedia Networks. *ACM Transactions on Information Systems*. 7, 1 (1989), pp. 58-84.

Date Due

Lib-26-67

MIT LIBRARIES DUPL

3 9080 00779668 0